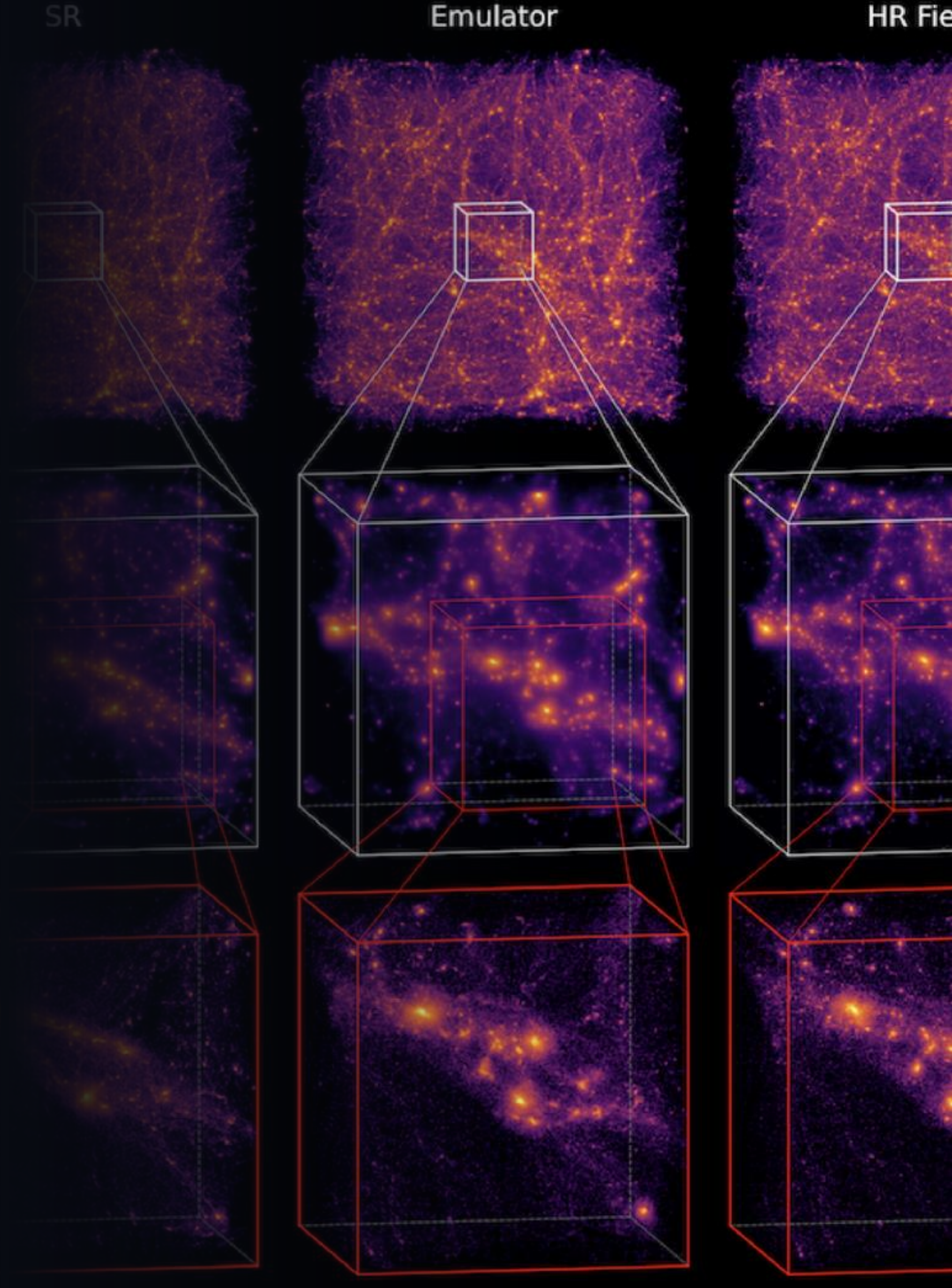
COSMIC HORIZONS CONFERENCE

AI Accelerators for Cosmological Simulations

Super-Resolution and Emulation — with an LLM application for simulations.

Xiaowen Zhang
Carnegie Mellon University

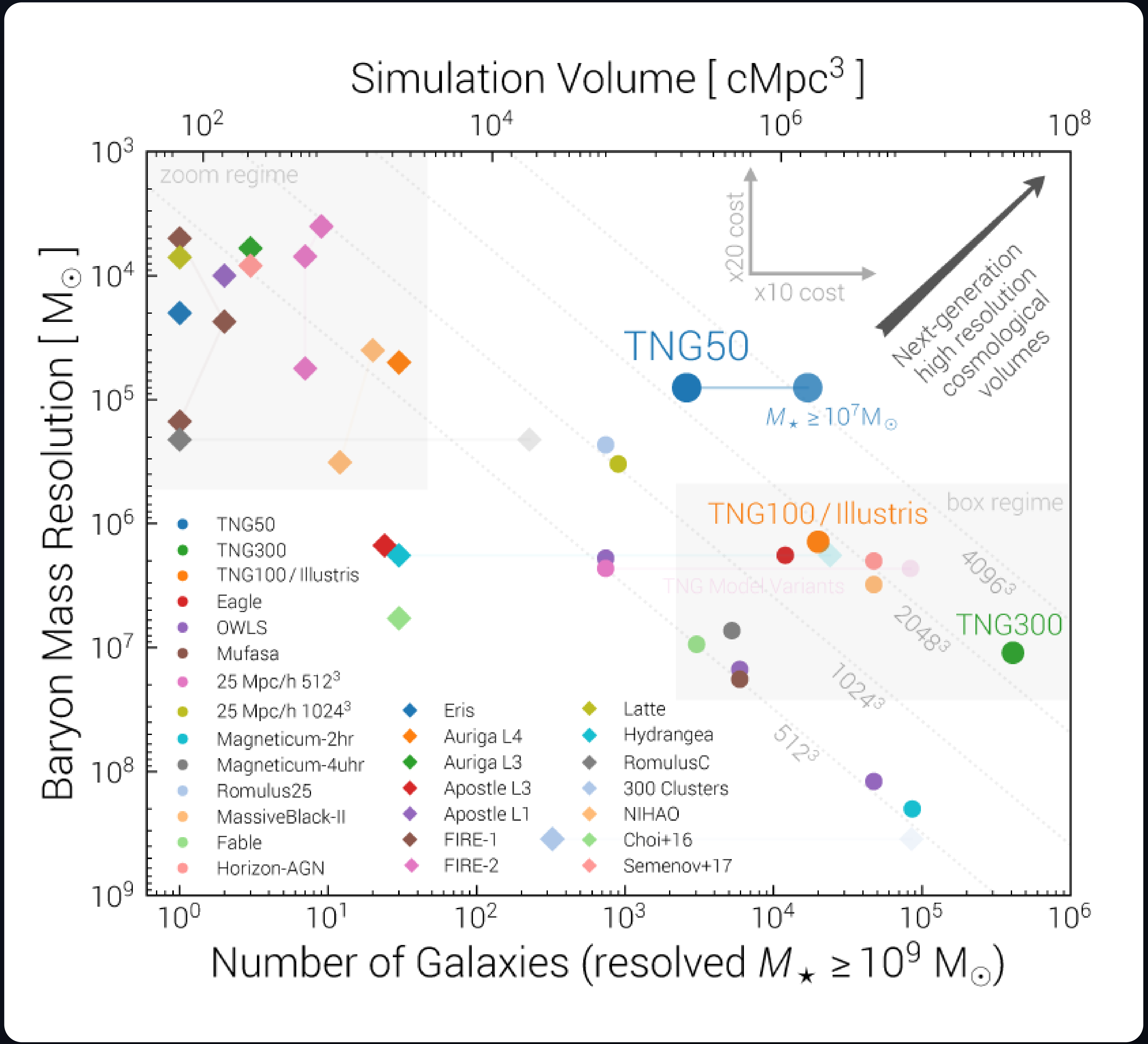


SUPER-RESOLUTION

Motivation

- Simulations are essential to interpret surveys and enable **simulation-based inference** (SBI).
- The goal: push to **larger volumes** and **higher mass resolution** to match upcoming data.
- The reality: state-of-the-art runs cost **millions of node-hours** .

FIXED COMPUTE **Volume** ⇄ **Resolution**



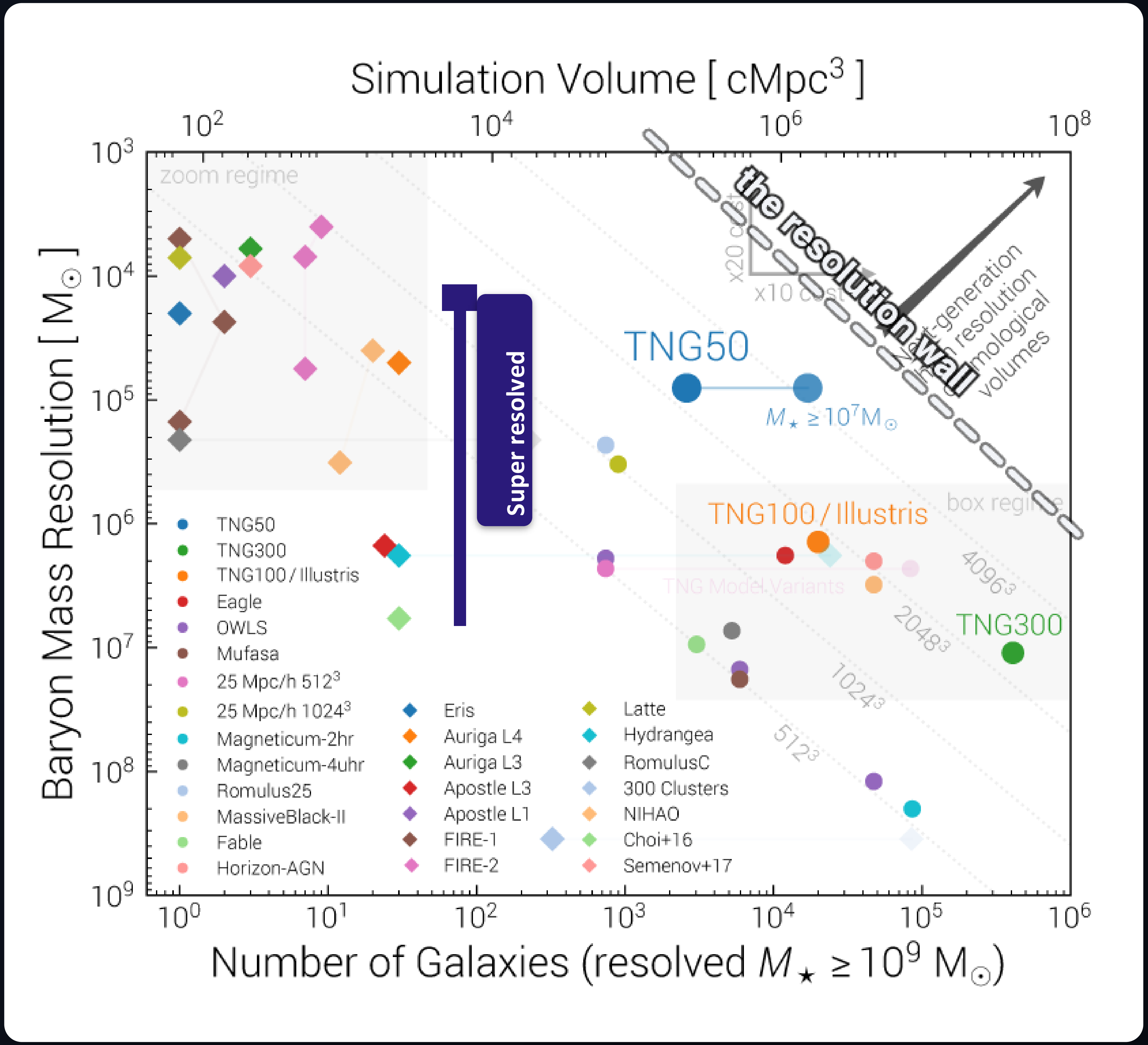
Mass resolution vs. volume · adapted from the TNG project

Source ·Base figure: IllustrisTNG project · SR method: Zhang et al. 2024, MNRAS 528, 281 (Paper III).

SUPER-RESOLUTION

Breaking the Volume–Resolution Wall

- Break the wall with **DL super-resolution**.
- Enhance mass resolution at fixed volume — learn a mapping that **adds small-scale structure** while preserving large-scale modes.
- Amortize cost: **train once, upscale many**.



↗ SR boost: shift toward higher resolution at the same volume

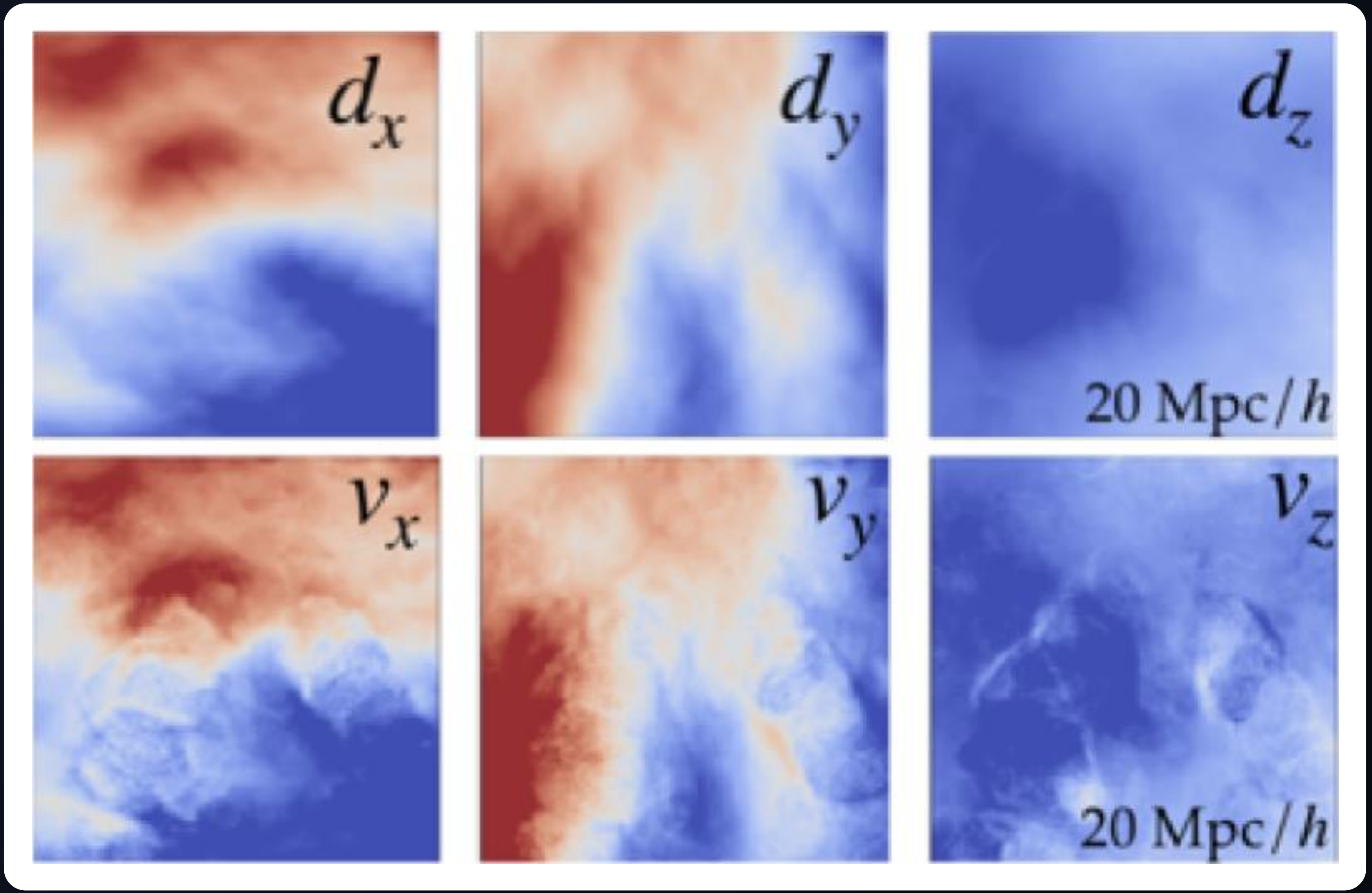
Source ·Base figure: IllustrisTNG project · SR method: Zhang et al. 2024, MNRAS 528, 281 (Paper III).

SUPER-RESOLUTION

Model Setup — Lagrangian Description

$d_i = x_i - q_i$ (i = 1 … N)

- **d** displacement · **x** final position · **q** initial uniform grid.
- Mass naturally conserved; particles tracked along trajectories.
- Output format matches a real simulation — 6D phase space (x, y, z, v_x, v_y, v_z).
- **Training data:** paired N-body simulations, LR 64^3 → HR 512^3, all in a fixed $100\ h^{-1}\mathrm{Mpc}$ box.



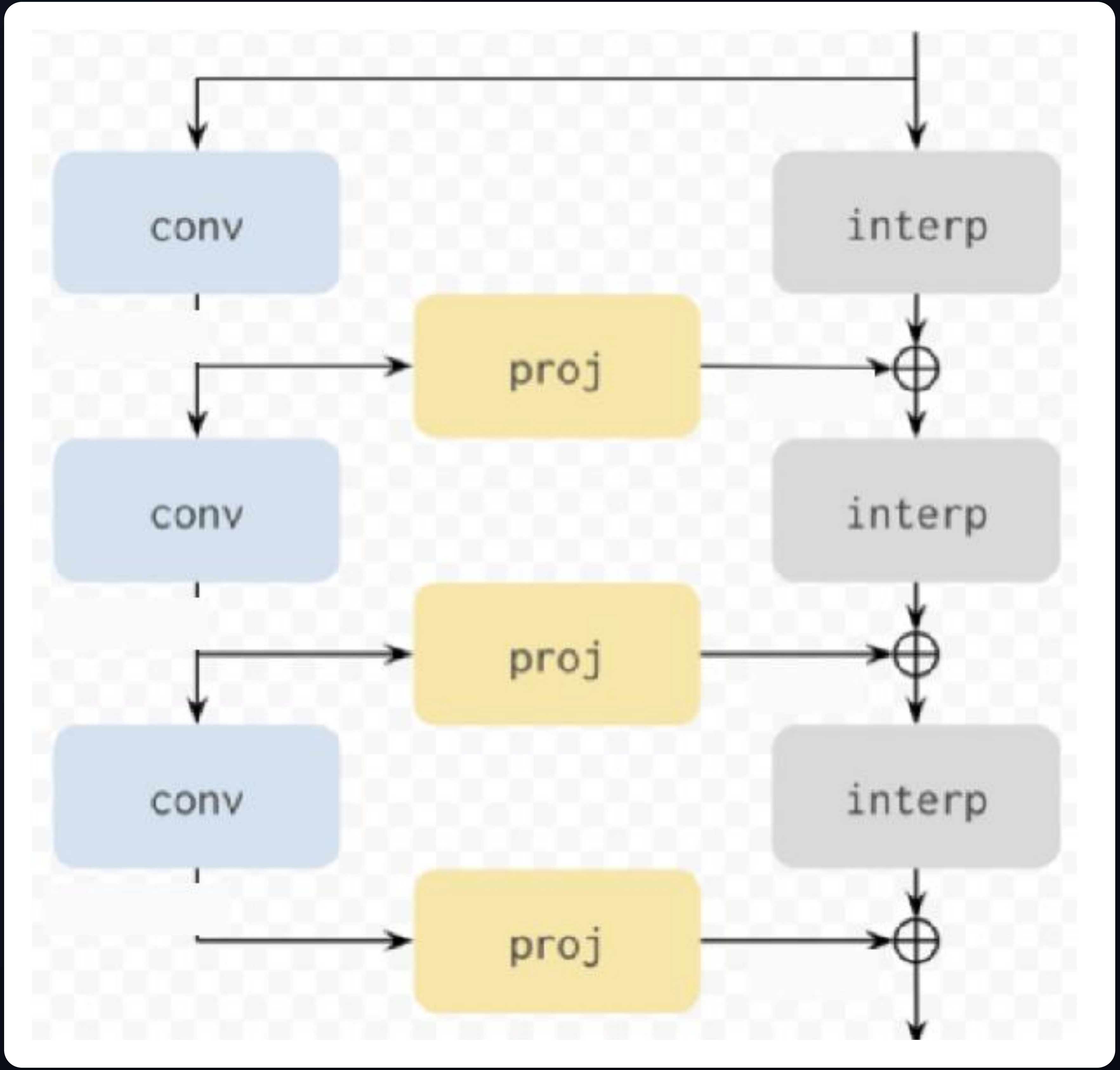
Displacement (d_x, d_y, d_z) and velocity (v_x, v_y, v_z) fields · 20 Mpc/h

Source · Zhang et al. 2024, MNRAS 528, 281 (Paper III).

SUPER-RESOLUTION

Conditional SR Model

- **Input:** LR Lagrangian particles + scale factor a (time conditioning).
- **Output:** SR 6D particle field — **512×** more particles (8× per side).
- **Training:** generator + critic, Wasserstein loss (Earth-mover distance).
- **Stochastic:** control small-scale randomness with Gaussian latent noise.
- **Fully convolutional → translationally equivariant:** split a large simulation into small chunks, super-resolve each, and stitch back.



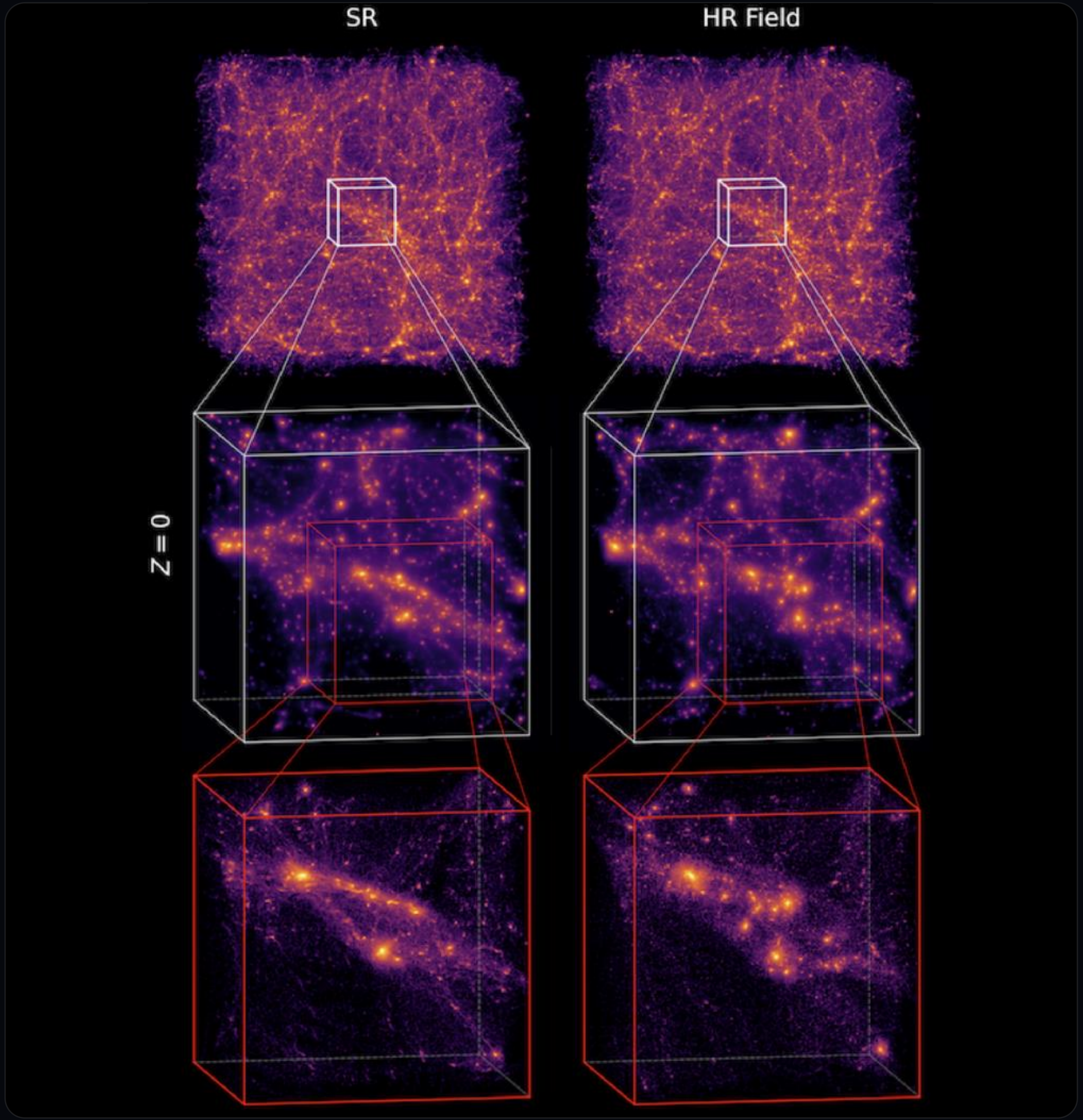
Progressive generator: conv → proj → interp across scales

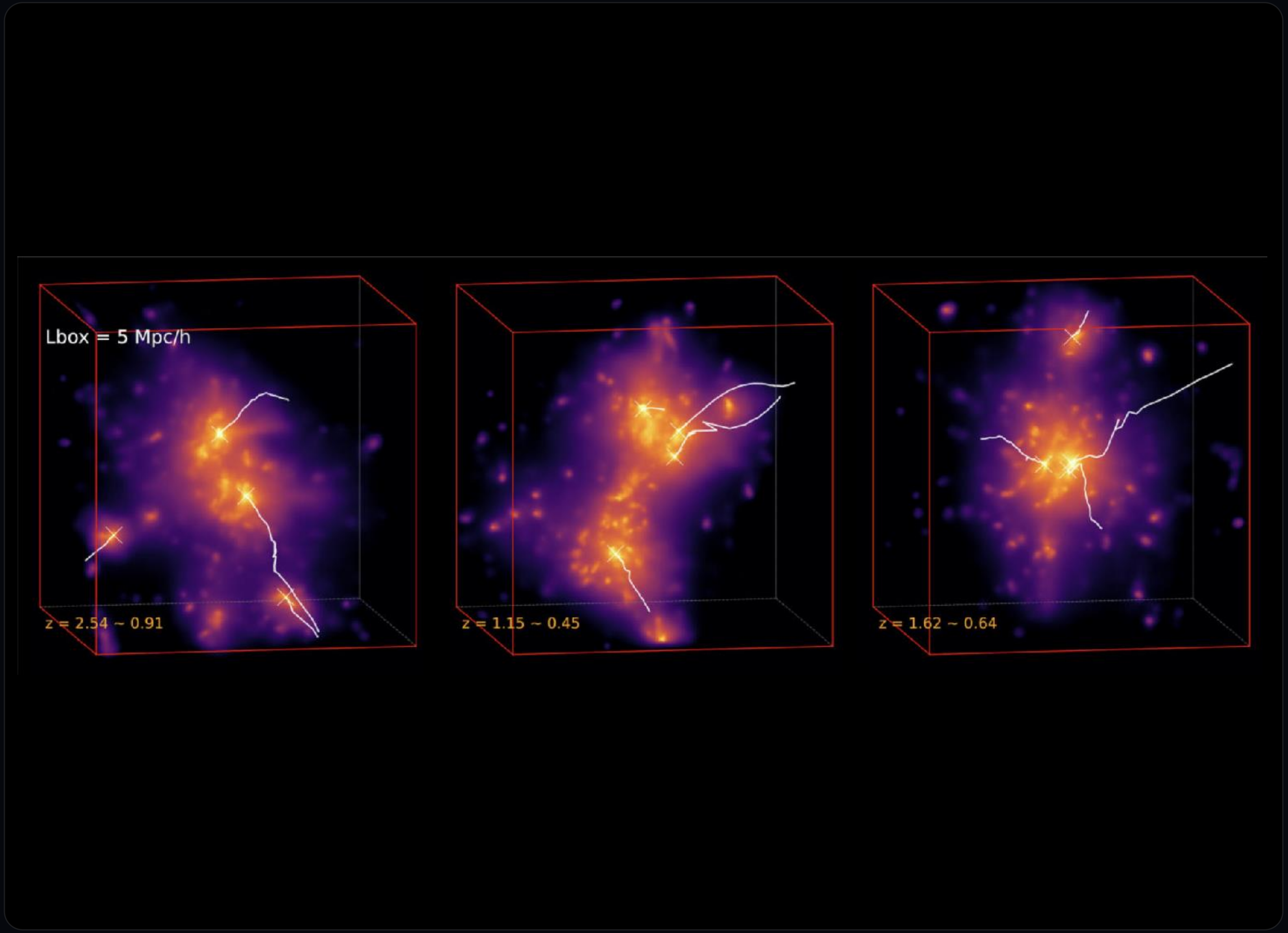
Source ·Zhang et al. 2024, MNRAS 528, 281 (Paper III).

SUPER-RESOLUTION

Visualizations

SR recovers HR-like small scales — same filaments, halos, and a consistent assembly history over time.



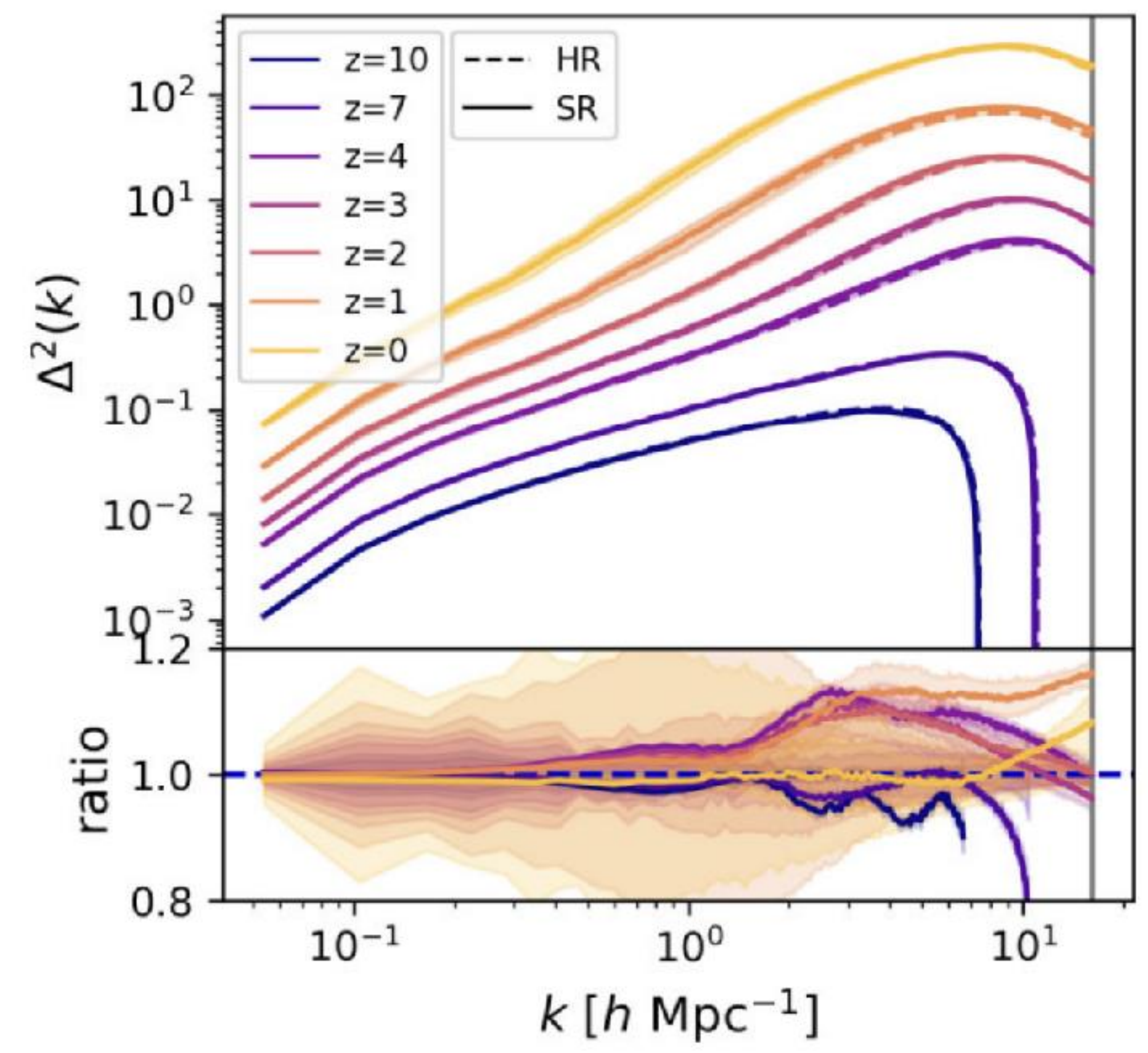


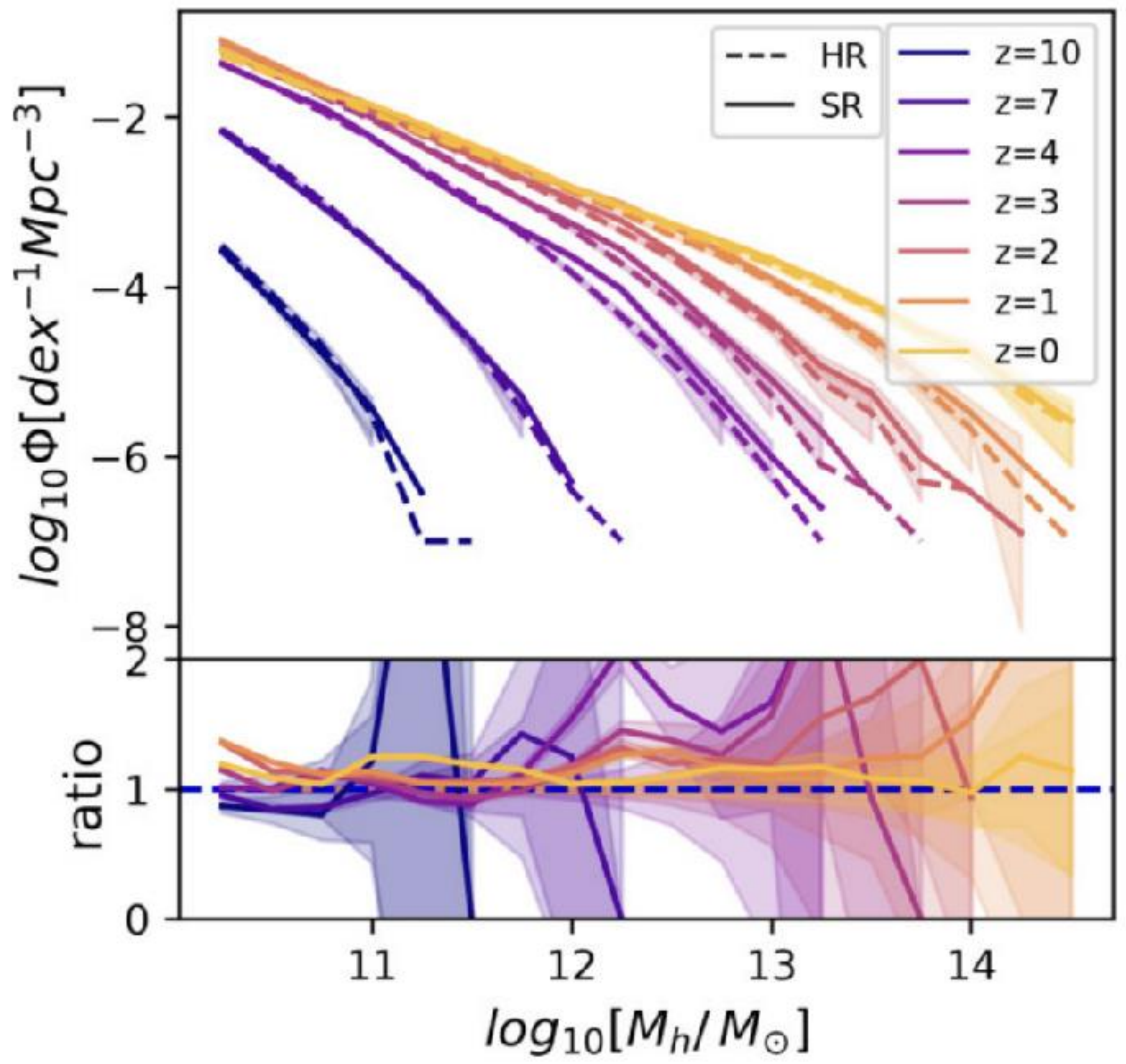
Particle trajectories around a massive halo (white tracks) — consistent assembly history across redshift intervals. $L_{box} = 5\ h^{-1}\mathrm{Mpc}$

Source · Zhang et al. 2024, MNRAS 528, 281 (Paper III).

SUPER-RESOLUTION

Statistics





Power spectra

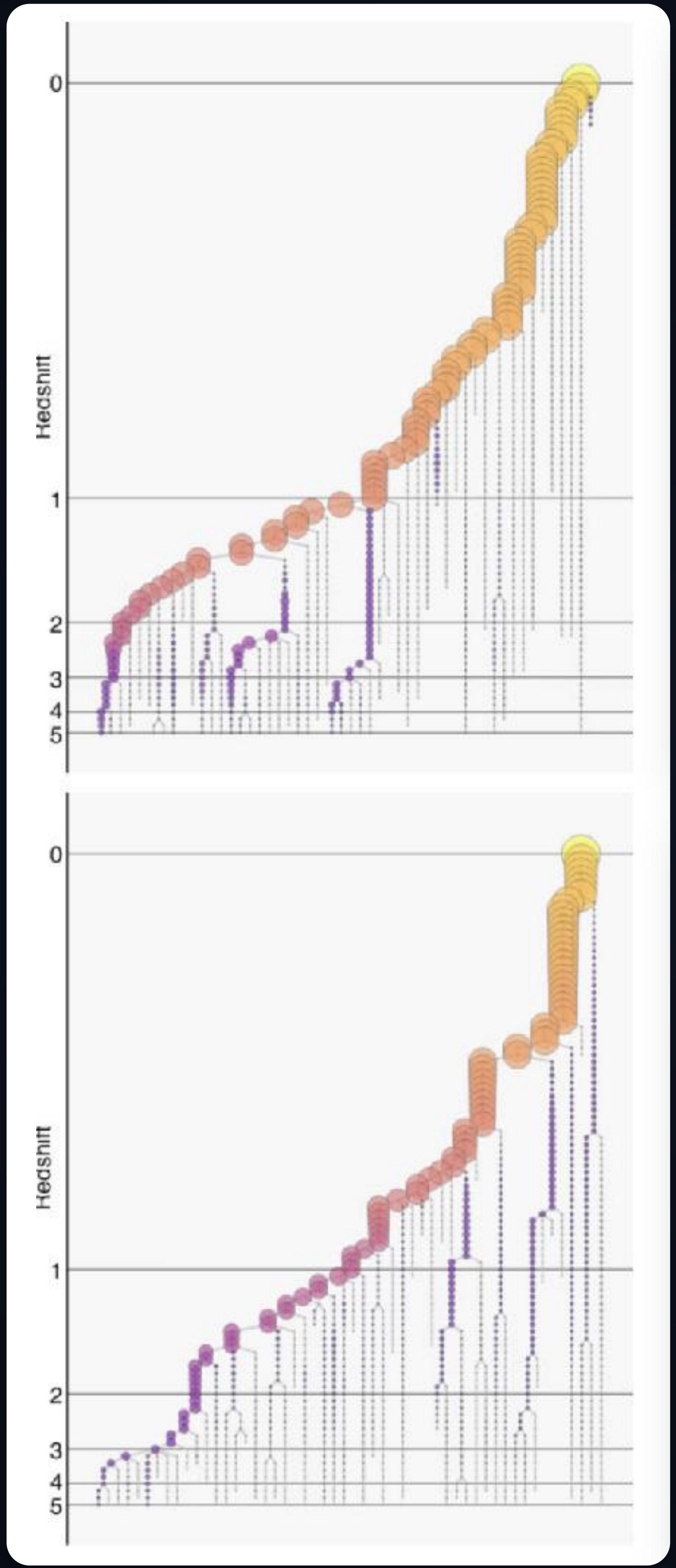
Halo mass function

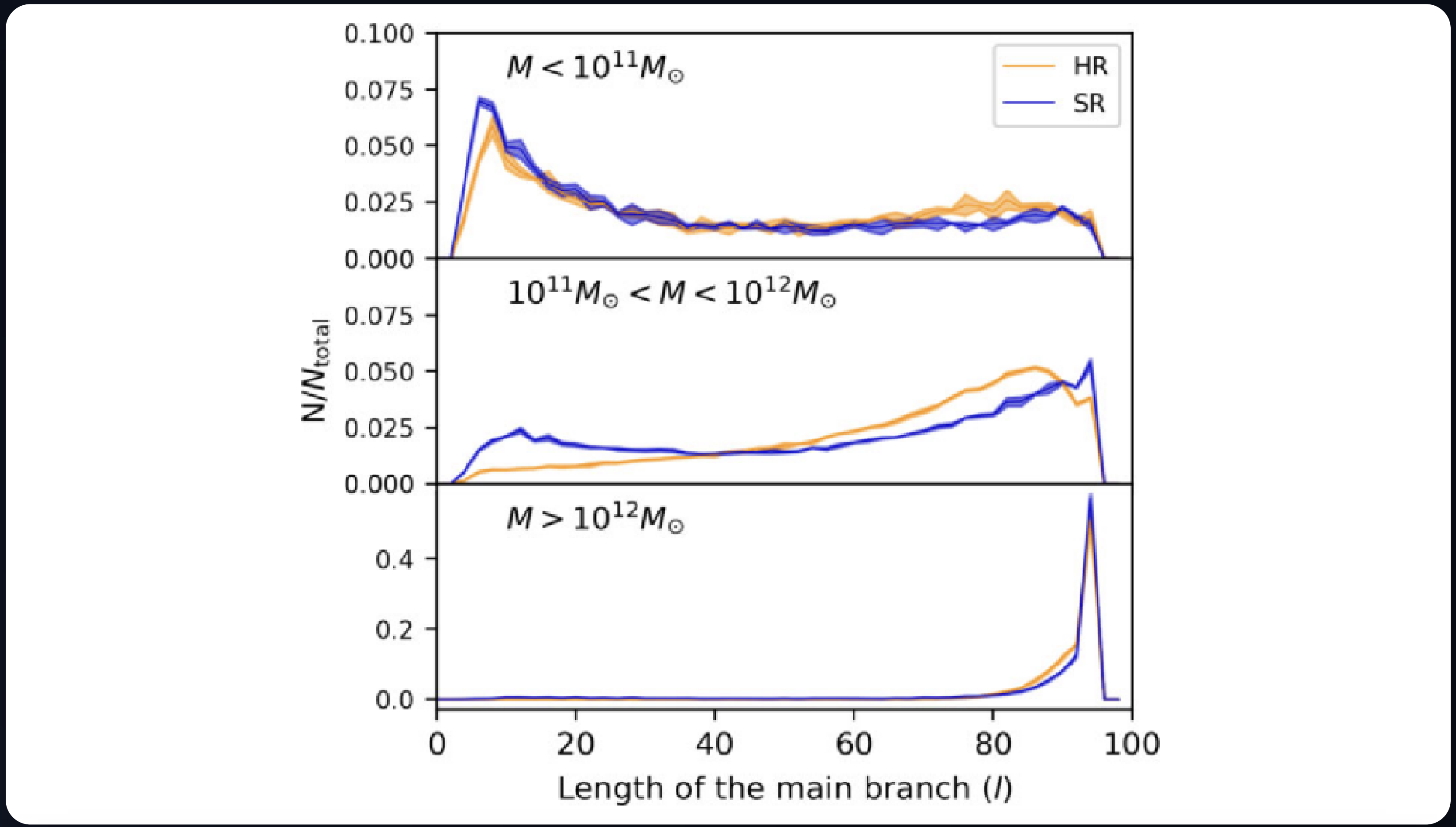
Source ·Zhang et al. 2024, MNRAS 528, 281 (Paper III).

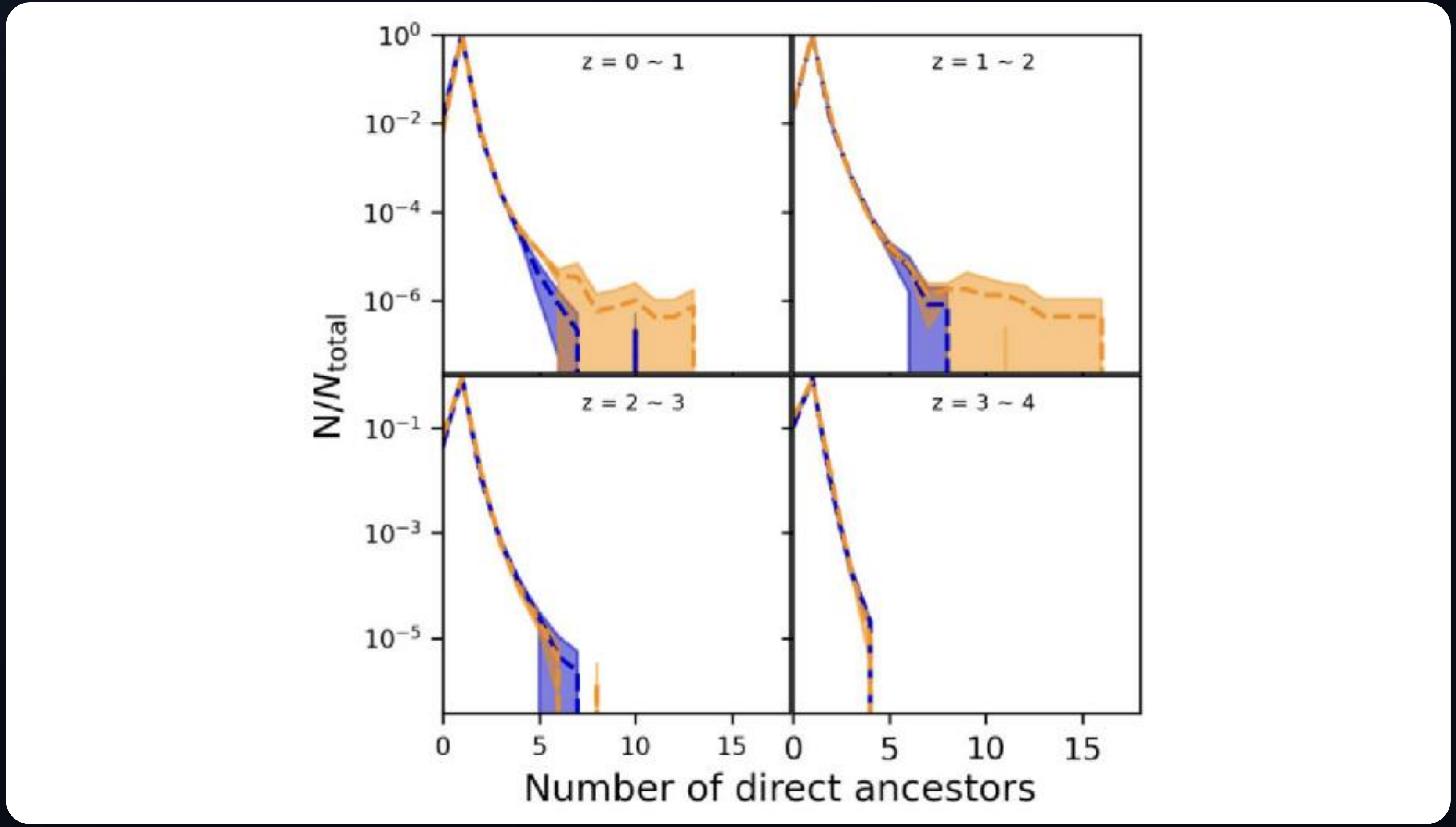
More Statistics — Merger Trees

- **Merger trees:** reproduce halo assembly — depth, branching, formation times.
- **Main-branch length:** distributions align across 3 mass bins.
- **Direct ancestors:** SR matches HR from z = 0 ~ 4.

100 snapshots · z = 5 → 0 · constant ∆log(a)







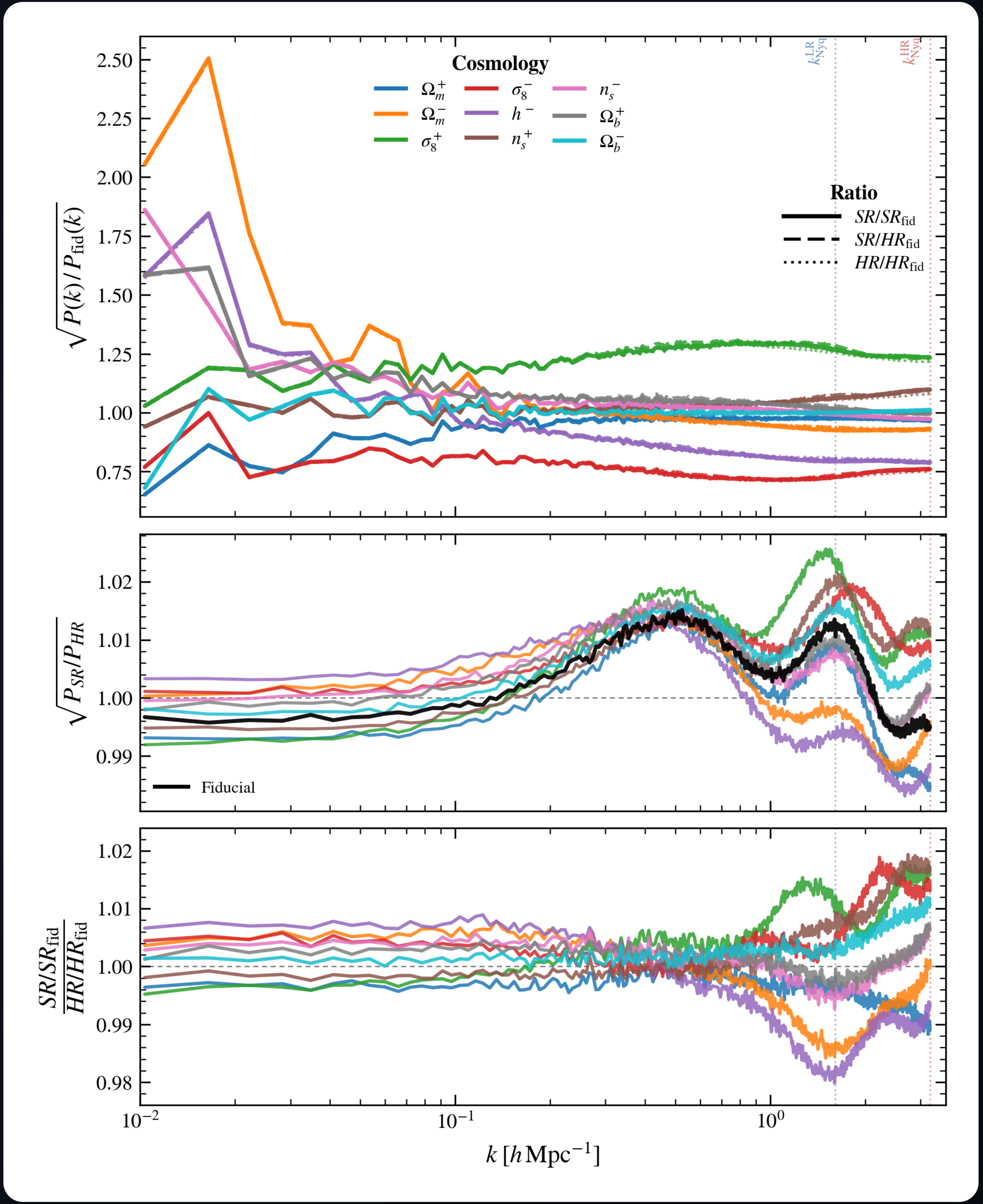
Source · Zhang et al. 2024, MNRAS 528, 281 (Paper III).

SUPER-RESOLUTION IN PROGRESS

Extending SR Across Cosmologies

- **Beyond a single cosmology:** train SR to generalize across a broad parameter grid.
- **Quijote suite:** samples Ω_m, σ_8, h, Ω_b, n_s.
- **Power spectrum:** SR reproduces the cosmology response, not just the fiducial run.

Enables simulation-based inference across parameter space.



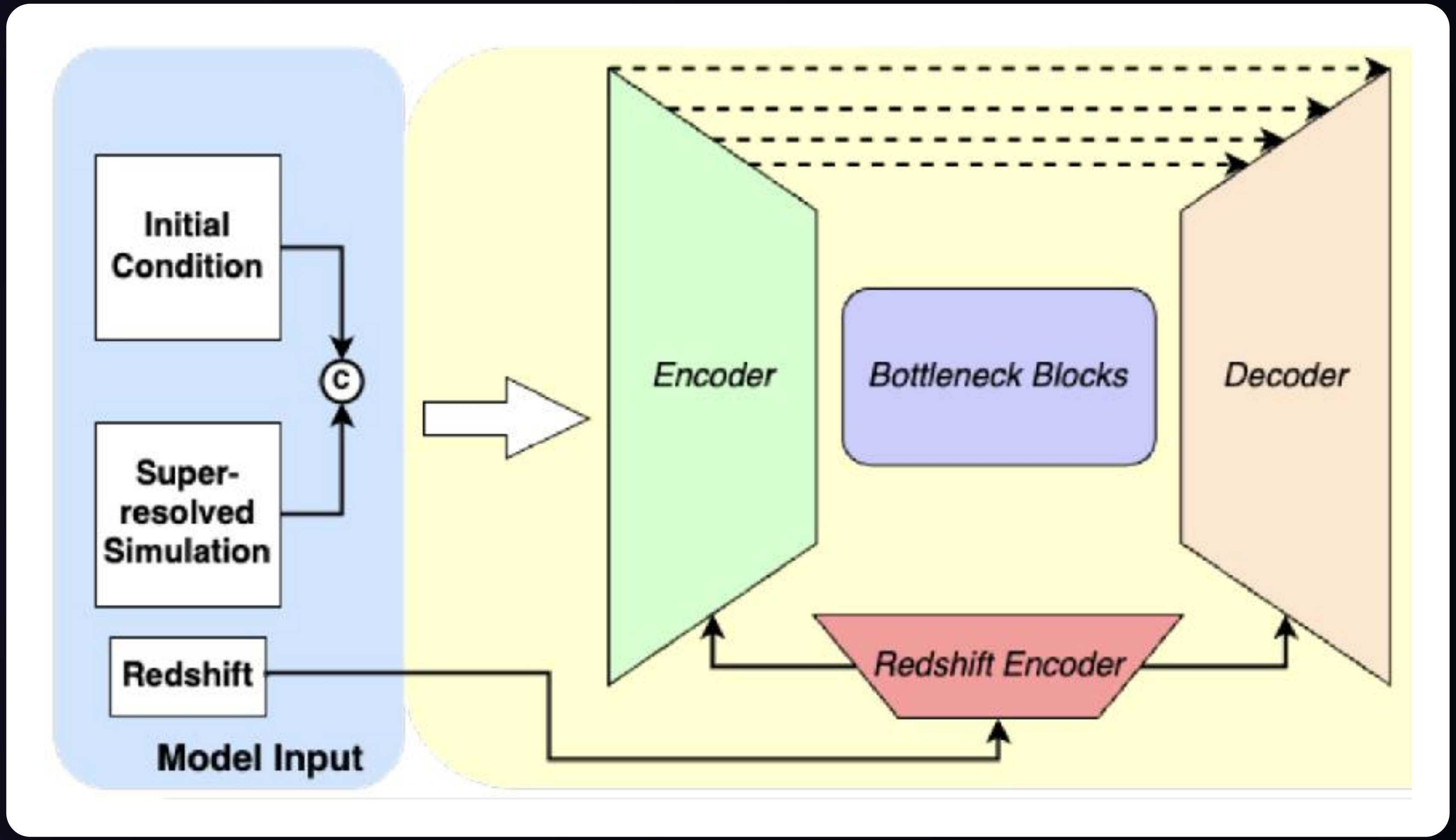
P(k) ratios across the cosmology grid · each line a different cosmology

Source ·Zhang et al., in prep. · Quijote simulation suite (Villaescusa-Navarro et al. 2020).

EMULATOR

From Stochastic to Deterministic

- **Challenge:** one LR → many plausible SR (stochastic ensemble).
- **Need:** the specific HR realization for given HR initial conditions (HRIC).
- **Goal:** without running HR — use LR + HRIC.
- **Approach:** transform SR output into the deterministic field matching HR with the same HRIC.



MULTI-TERM OBJECTIVE

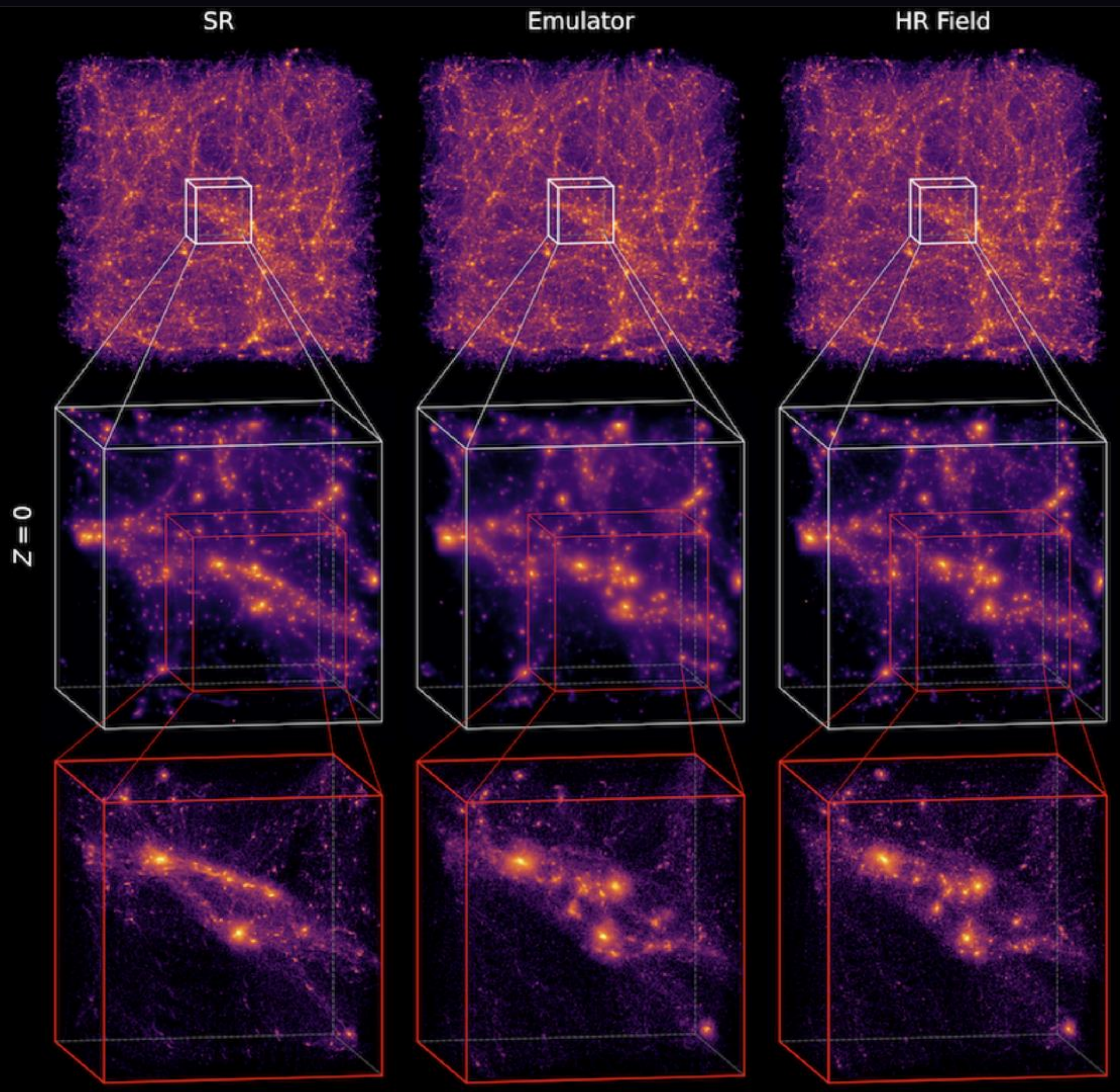
- **Lagrangian fields** — displacement & velocity.
- **Eulerian density**
- **Distribution-matching**

Source ·Zhang et al. 2025, Open J. Astrophys. 8 (Paper IV).

EMULATOR

Visualizations

The emulator aligns small-scale structure to HR under the same initial conditions — SR, emulator, and true HR field side by side.



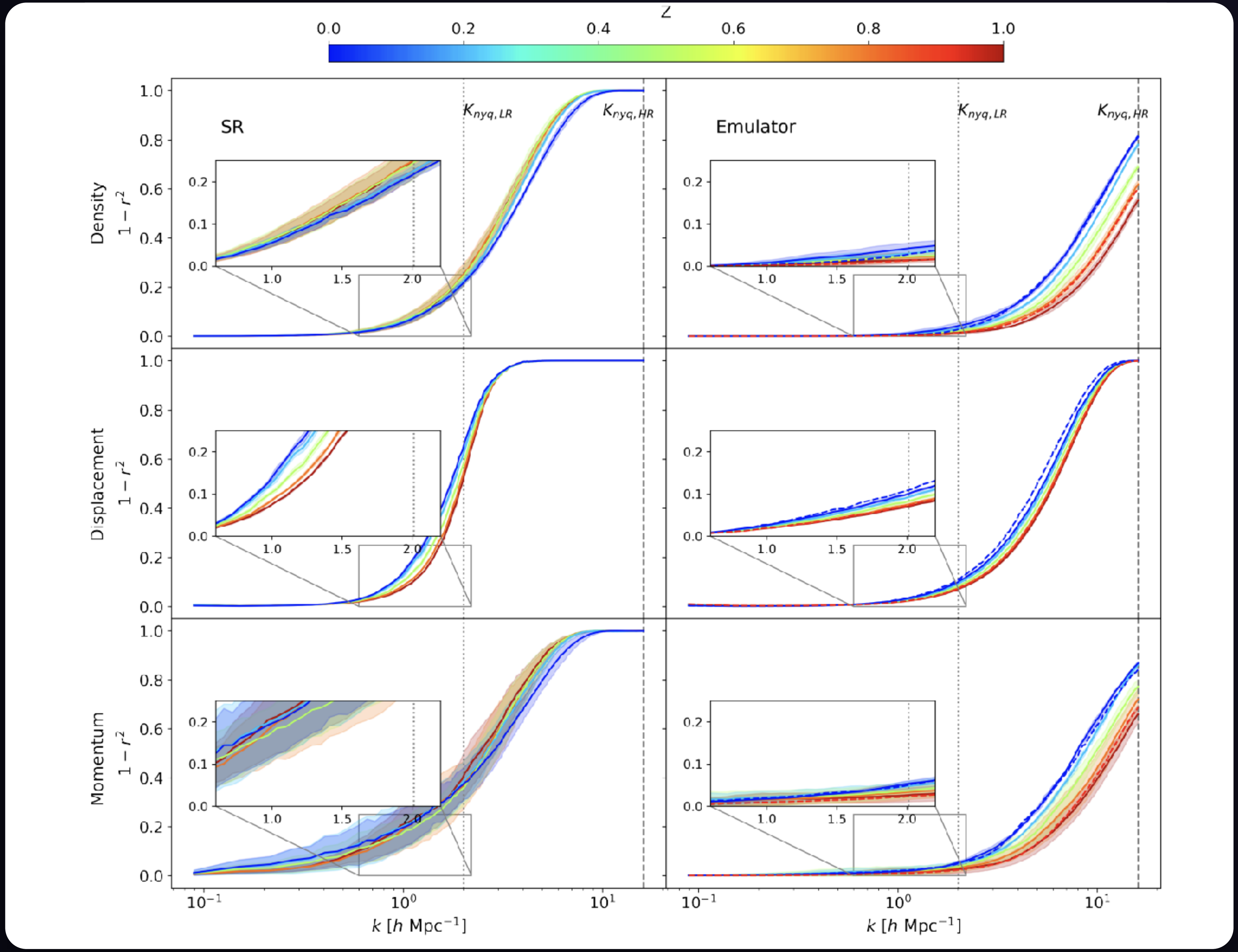
Source ·Zhang et al. 2025, Open J. Astrophys. 8 (Paper IV).

EMULATOR

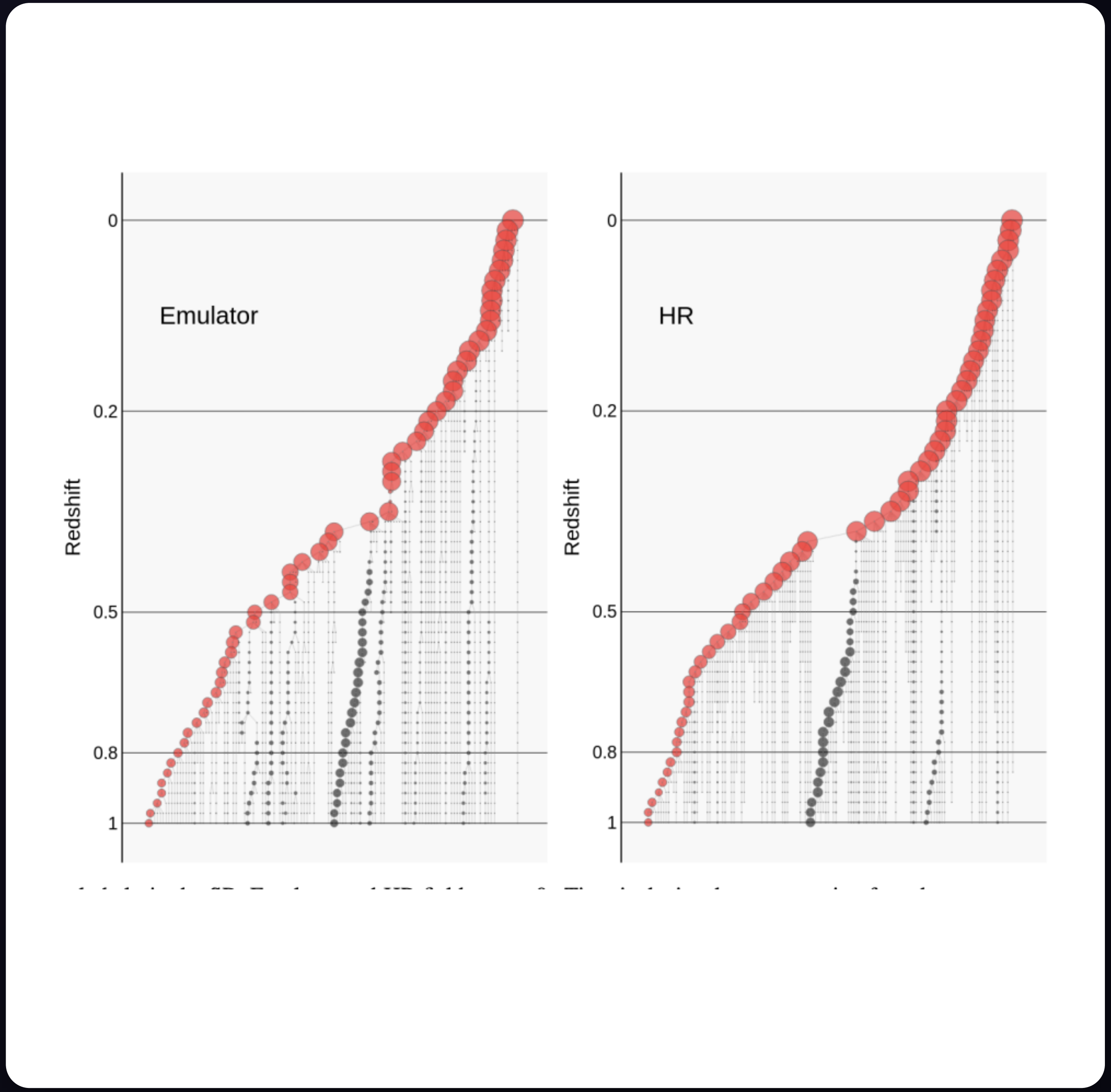
Fourier Cross-Correlation & Merger Trees

60 snapshots · z = 1 → 0 · constant ∆log(a)

- **Fourier cross-correlation:** mode-by-mode agreement with HR stays high to small scales, across redshift ($L_{box} = 10$ h^{-1}Mpc).
- **Merger trees:** almost identical merger history.



$$r(k) = \frac{\langle \delta_{\rm pred}\, \delta_{\rm true} \rangle}{\sqrt{\langle \delta^2_{\rm pred} \rangle \langle \delta^2_{\rm true} \rangle}}$$

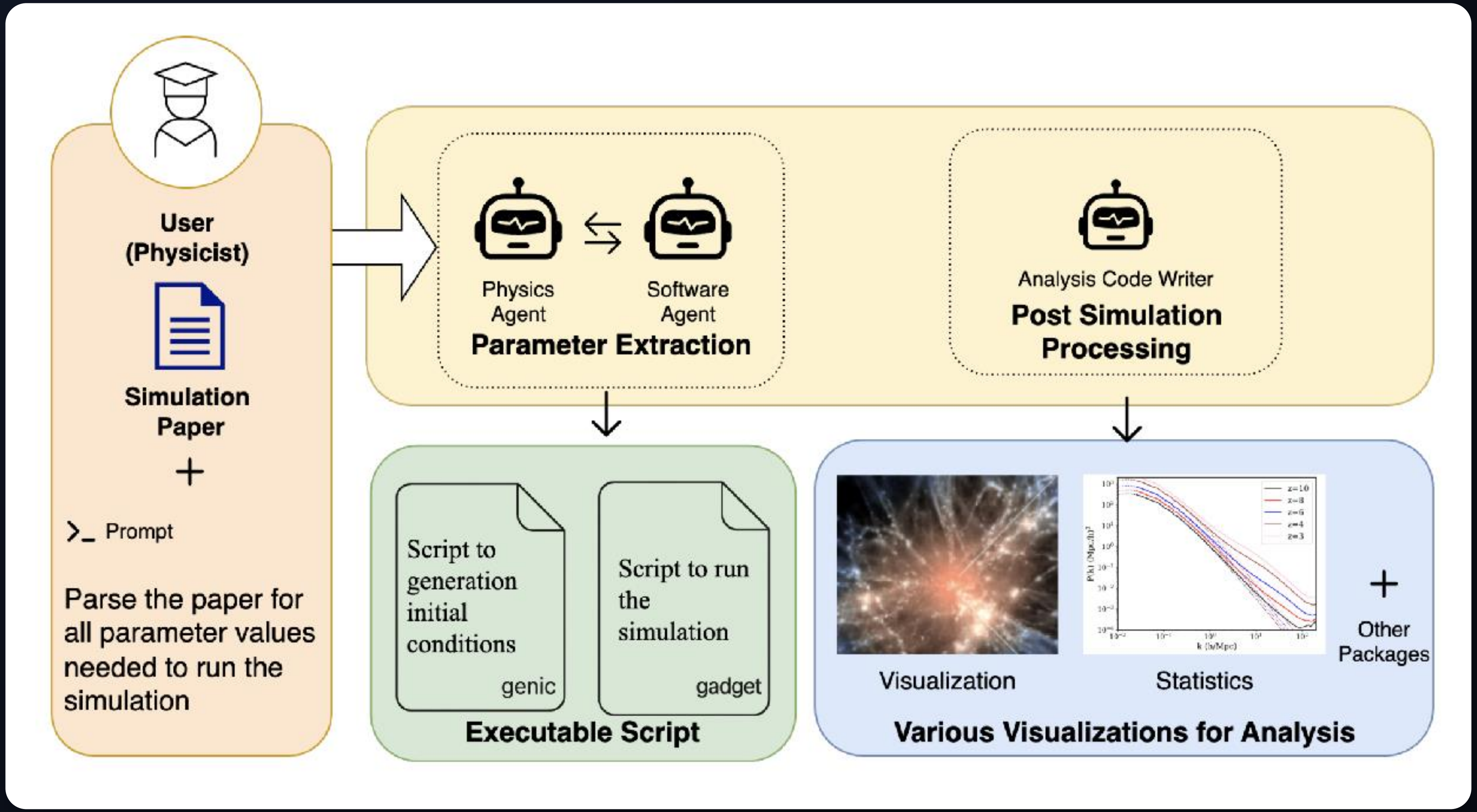


Source ·Zhang et al. 2025, Open J. Astrophys. 8 (Paper IV).

MULTI-AGENT SYSTEM

A "Trustworthy" LLM Tool

- **Accelerates the workflow** — automates setup, not "doing science."
- **Three specialized agents:** Physics Agent reads the paper, Software Agent enforces the manual, Analysis Code Writer plots results.
- **Iterative discussion** → parameters that are physically meaningful, consistent, and software-compliant.
- **Closes the loop:** paper → parameter file → simulation → preliminary analysis.



Source · Zhang et al. 2025, SimAgents, IJCNLP-AACL 2025 (System Demo., pp. 55–66).

MULTI-AGENT SYSTEM

How Good Is It?

BASELINES VS. OURS

- **CoT** — one agent, chain-of-thought reasoning.
- **EoT** — two agents exchanging thoughts.
- **Ours** — two agents: a generator + a verifier.

ERROR METRICS — AVG. PER SIMULATION (LOWER = BETTER)

- **Value error** — parameter exists, but its value or unit is wrong.
- **Type error** — parameter from an incompatible sim context (e.g. gas settings in a dark-matter-only run).
- **Hallucination** — invents parameters absent from the software manual.

METHOD	MICRO-F1	PRECISION	RECALL
CoT (1-Agent)	93.64	92.46	94.84
EoT (2-Agent)	94.95	93.87	96.05
Ours (2-Agent)	**98.67**	**97.80**	**99.55**

METHOD	VALUE ERR ↓	TYPE ERR ↓	HALLUC. ↓
CoT (1-Agent)	0.97	0.51	**0.21**
EoT (2-Agent)	1.21	0.21	0.34
Ours (2-Agent)	**0.46**	**0.02**	**0.30**

Source ·Zhang et al. 2025, SimAgents, IJCNLP-AACL 2025 (System Demo., pp. 55–66) · human evaluation, GPT-4 backbone.

WRAP-UP

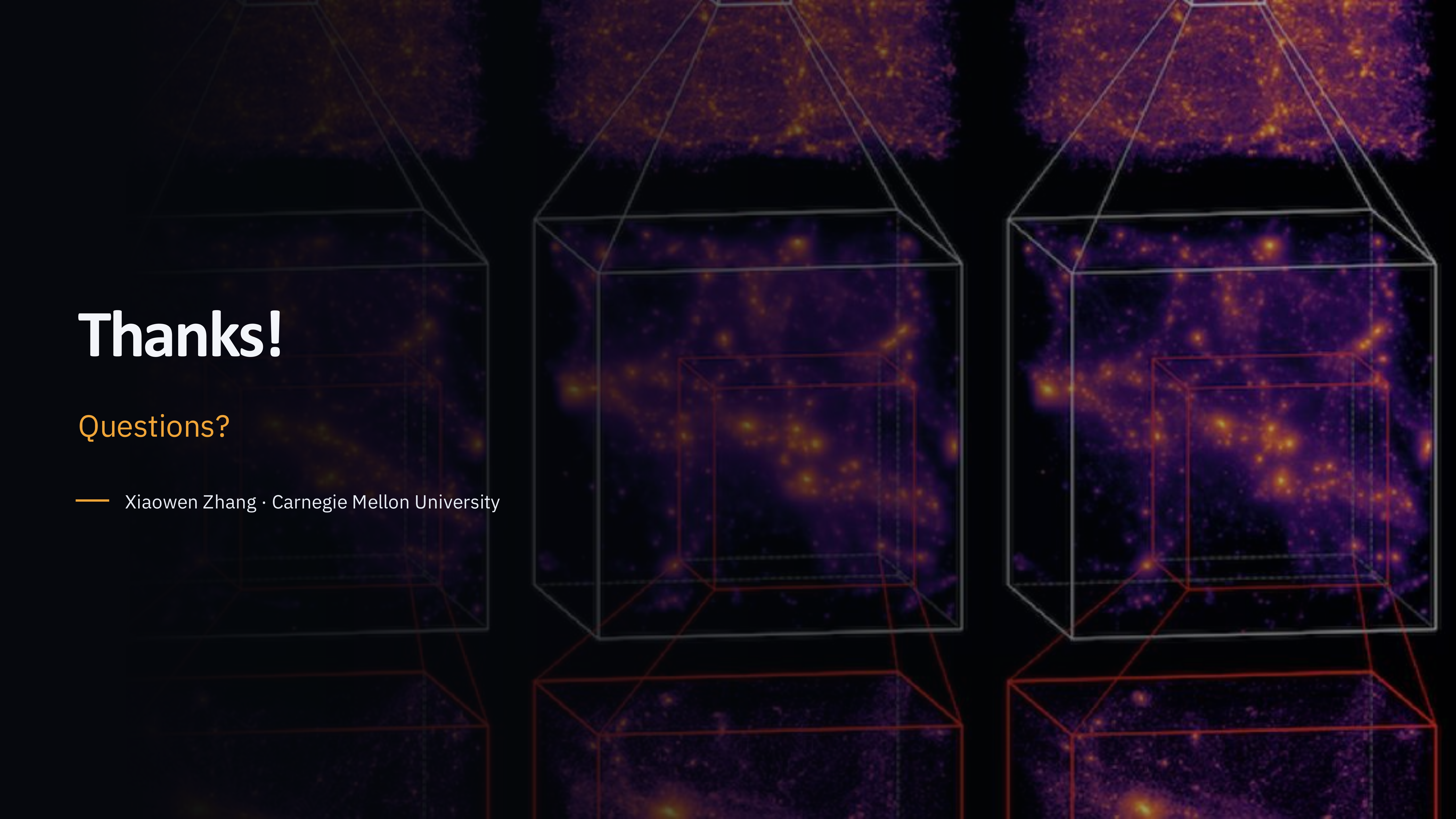
Summary & Outlook

WHAT WE BUILT

- **SR model** — full 6D phase-space output at **512×** higher mass resolution across redshift.
- **Emulator** — adjust SR output to match a specific HR simulation given its ICs.
- **Multi-agent tools** — accelerate the simulation setup workflow.

NEXT STEPS

- **Cosmology coverage**
- **More physics** — extend to hydro, cooling, feedback, black-hole models.



Thanks!

Questions?

Xiaowen Zhang · Carnegie Mellon University